Empirical Output Distribution of Good Delay-Limited Codes for Quasi-Static Fading Channels

Silas L. Fong and Vincent Y. F. Tan

Abstract

This paper considers delay-limited communication over quasi-static fading channels under a long-term power constraint. A sequence of length-n delay-limited codes for a quasi-static fading channel is said to be *capacity-achieving* if the codes achieve the *delay-limited capacity*, which is defined to be the maximum rate achievable by delay-limited codes. The delay-limited capacity is sometimes referred to as the *zero-outage capacity* in wireless communications. The delay-limited capacity is the appropriate choice of performance measure for delay-sensitive applications such as voice and video over fading channels. It is shown that for any sequence of capacity-achieving delay-limited codes with vanishing error probabilities, the normalized relative entropy between the output distribution induced by the length-n code and the n-fold product of the capacity-achieving output distribution, denoted by $\frac{1}{n}D(p_{Y^n}\|p^*_{Y^n})$, converges to zero. Additionally, we extend our convergence result to capacity-achieving delay-limited codes with non-vanishing error probabilities.

I. Introduction

In information and coding theory, the search for good codes for various classes of channels is of paramount importance. By "good" we mean that the code is *reliable*, i.e., its (average or maximal) probability of error is arbitrarily small when the blocklength is sufficiently large. In addition to being good, communication engineers would also like the code to be *optimal* in the sense that the rate of the code (the ratio of the logarithm of the number of codewords to the blocklength) converges to the channel capacity as the blocklength grows. The search for optimal good codes for memoryless channels, however, is known to be challenging and has remained elusive for decades.

As a result, information theorists have resorted to characterizing the nature or properties of good codes that are asymptotically optimal. One of the useful characterizations is in terms of the so-called approximation of output statistics [1], studied by Han and Verdú. In [1, Theorem 15], general channels [2] that satisfy the strong converse property and whose input alphabets are finite were considered. For this class of channels, it was shown that for reliable codes whose rates approach the channel capacity, the normalized relative entropy between p_{Y^n}, the output distribution induced by the length-n code, and $p^*_{Y^n}$, the n-fold product of the (unique) capacity-achieving output distribution, converges to zero, i.e.,

$$\lim_{n\to\infty}\frac{1}{n}D(p_{Y^n}\|p^*_{Y^n}) = 0. \tag{1}$$

This result implies that good capacity-achieving codes must necessarily be such that its empirical output distribution is close to the capacity-achieving output distribution in the sense of (1). Thus to find optimal codes, communication engineers can and indeed must restrict their search to this class of codes.

The seminal work by Han and Verdú [1] was subsequently generalized by Shamai and Verdú [3] who lifted the restriction concerning the finiteness of the input alphabet of the channel. They showed that (1) holds under the condition that the capacity of the general channel can be written as

$$C = \lim_{n\to\infty}\sup_{p_{X^n}}\frac{1}{n}I(X^n;Y^n). \tag{2}$$

Indeed, when the input alphabet is finite and the strong converse property holds, the capacity [1, Theorem 8] is given by the expression in (2).

In yet another generalization, Polyanskiy and Verdú [4] studied the properties of ε-good codes under the *maximal* probability of error formalism and with deterministic encoders. These are codes whose maximal probabilities of

Silas L. Fong and Vincent Y. F. Tan are with the Department of Electrical and Computer Engineering, National University of Singapore (NUS), Singapore (e-mail: {`silas_fong,vtan`}`@nus.edu.sg`). Vincent Y. F. Tan is also with the Department of Mathematics, NUS.

error are bounded above by a non-vanishing constant $\varepsilon \in (0,1)$. The study of ε-good codes has gained prominence recently due to the interest in the finite blocklength regime [5] and second-order asymptotics [5]–[7]. Polyanskiy and Verdú showed that despite this generalization, the approximation in (1) continues to hold for a large class of channels including discrete memoryless channels (DMCs) [4, Theorems 6 & 7] and additive white Gaussian noise (AWGN) channels under the peak power constraint [4, Theorem 8], i.e., that every transmitted codeword $(x_1, x_2, \ldots, x_n)$ must satisfy

$$\frac{1}{n}\sum_{k=1}^{n} x_k^2 \leq P \tag{3}$$

for some power $P > 0$.

A. Motivation

For communication over wireless fading channels [8], it is also of interest to study codebooks whose cost constraints are in the *average-over-the-codebook-and-fading* sense, i.e.,

$$\frac{1}{M_n}\sum_{w=1}^{M_n}\left(\frac{1}{n}\sum_{k=1}^{n} \mathrm{E}\left[X_k^2(w)\right]\right) \leq P \tag{4}$$

where M_n denotes the codebook size of the length-n code and the expectation is taken over the fading statistics. This is known in the wireless communication community as the *long-term power constraint* [8,9]. This more relaxed constraint is useful and practical in wireless communication as it allows for the dynamic allocation of available power based on the current fading state of the channel.

This work investigates a class of wireless fading channels called *quasi-static fading channels* [10, Section 5.4.1], where the fading coefficient is random but remains constant during the course of transmission. As explained in [10, Section 5.4.1], the quasi-static fading channel can be used to model slow fading (time-invariant) channels where the delay requirement is short compared to the channel coherence time. We are interested in the *delay-limited* capacity [11] (also called the *zero-outage* capacity [12, Section 4.2.4]) of the quasi-static fading channel, which is the maximum achievable rate under the assumption that the maximal error probability over all non-zero fading coefficients vanishes as the blocklength grows. As explained in [11, Section I], the delay-limited capacity is the appropriate choice of performance measure for delay-sensitive applications such as voice and video that could not tolerate long delays. Since the quasi-static fading channel and the corresponding delay-limited capacity are generalizations of the AWGN channel and the corresponding capacity, we are motivated to study whether the property in (1) continues to hold for good delay-limited codes under the power constraint in (4).

B. Main Contribution

The main contribution of this work is to demonstrate that the convergence result in (1) continues to hold for good *delay-limited* codes for the quasi-static fading channel. A crucial step to proving the convergence result is to establish the convergence of the average signal power received at the destination for good delay-limited codes as stated in Lemma 3. Proving that the average received signal power converges to its appropriate limit is non-trivial and rather technical. In particular, we need to combine several well-known facts such as continuity of measure as well as the monotonicity of a certain function.

In addition, we extend our convergence result for good delay-limited codes to ε-good delay-limited codes, i.e., capacity-achieving codes with non-vanishing error probabilities. The extension of the convergence result is performed by combining the aforementioned proof techniques with an existing non-asymptotic multi-letter converse bound in [4] for the AWGN non-fading channel under the peak power constraint (3). The convergence property in (1) for the AWGN non-fading channel (cf. [4, Theorem 8]) can be viewed as a special case of our convergence results for the quasi-static fading channel when there is only one fading state.

C. Paper Outline

This paper is organized as follows. The next subsection presents the notation used in this paper. Section II provides the problem formulation of the quasi-static fading channel under a long-term power constraint and presents our

main result — the convergence property in (1) for good delay-limited codes. Section III presents the proof of the convergence result. Section IV extends our convergence result for good delay-limited codes to ε-good delay-limited codes.

D. Notation

We use $\Pr\{\mathcal{E}\}$ to represent the probability of the event $\mathcal{E}$, and we let $\mathbf{1}\{\mathcal{E}\}$ be the characteristic function of $\mathcal{E}$. We use a capital letter X to denote an arbitrary random variable with alphabet $\mathcal{X}$, and use the small letter x to denote a realization of X. We use X^n to denote a random tuple $(X_1, X_2, \ldots, X_n)$, where the components X_k have the same alphabet $\mathcal{X}$.

The following notations are used for any arbitrary random variables X and Y. We let p_X and $p_{Y|X}$ denote the probability distribution of X and the conditional probability distribution of Y given X respectively. We let $\Pr_{p_X}\{g(X) \in \mathcal{A}\}$ denote $\int_{x\in\mathcal{A}} p_X(x)\mathbf{1}\{g(x) \in \mathcal{A}\}\,\mathrm{d}x$ for any real-valued and any real-valued mapping g whose domain includes $\mathcal{X}$, and the expectation of $g(X)$ is denoted as $\mathrm{E}_{p_X}[g(X)]$. We let $p_X p_{Y|X}$ denote the joint distribution of (X, Y), i.e., $p_X p_{Y|X}(x, y) = p_X(x)p_{Y|X}(y|x)$ for all x and y. We let $\mathcal{N}(\cdot\,; \mu, \sigma^2) : \mathbb{R} \to [0, \infty)$ be the probability density function of a Gaussian random variable denoted by Z whose mean and variance are μ and σ^2 respectively such that

$$\mathcal{N}(z; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(z-\mu)^2}{2\sigma^2}}. \tag{5}$$

Similarly, we let $\mathcal{N}(\cdot\,; \mu, \sigma^2) : \mathbb{R}^n \to [0, \infty)$ be the joint probability density function of n independent copies of $Z \sim \mathcal{N}(z; \mu, \sigma^2)$ such that

$$\mathcal{N}(z^n; \mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}}\, e^{-\sum_{k=1}^{n}\frac{(z_k-\mu)^2}{2\sigma^2}}. \tag{6}$$

We will take all logarithms to base 2 throughout this paper, so all information quantities have units of bits. The sets of natural, real and non-negative real numbers are denoted by $\mathbb{N}$, $\mathbb{R}$ and $\mathbb{R}^+$ respectively. The Euclidean norm of a tuple $x^n \in \mathbb{R}^n$ is denoted by $\|x^n\| \triangleq \sqrt{\sum_{k=1}^{n} x_k^2}$.

II. Quasi-Static Fading Channel under a Long-Term Power Constraint

A. Problem Formulation and Main Result

We consider an additive white Gaussian noise (AWGN) fading channel that consists of one source and one destination, denoted by s and d respectively. The fading process is characterized by a random variable H (we follow the notation in [11, Section II] and use the upper-case letter H to denote the random fading coefficient). The fading process is assumed to be quasi-static, i.e., the fading coefficient H is selected randomly and kept constant during the course of transmission. The fading coefficient is assumed to be real and non-negative, and we let p_H denote its distribution. In addition, we assume $0 < \mathrm{E}_{p_H}\left[\frac{1}{H}\right] < \infty$, i.e.,

$$0 < \int_{\mathbb{R}^+} \frac{p_H(h)}{h}\,\mathrm{d}h < \infty\,, \tag{7}$$

which is a common assumption for fading channels with positive *zero-outage capacity* [12, Section 4.2.4]. Source s has the knowledge of H but destination d does not. In other words, we assume the availability of the channel state information at transmitter (CSIT).

Node s chooses message W from the set

$$\mathcal{W} \triangleq \{1, 2, \ldots, M_n\} \tag{8}$$

and sends W to node d in n time slots, where $M_n = |\mathcal{W}|$ denotes the message size and W is uniformly distributed over $\mathcal{W}$ and independent of H. The encoder at s is allowed to adapt to the fading coefficient H, and the encoding function for $H = h$ is denoted by $f_h : \mathcal{W} \to \mathbb{R}^n$ where $f_h(W)$ is the codeword corresponding to W when $H = h$. The set of encoding functions $\{f_h \,|\, h \geq 0\}$ satisfies the *long-term power constraint* [8,9]

$$\frac{1}{M_n}\sum_{w\in\mathcal{W}} \mathrm{E}_{p_H}\left[\|f_H(w)\|^2\right] \leq nP \tag{9}$$

for some fixed $P > 0$. This long-term power constraint is indeed an average power constraint where the average is taken over the realizations of both the fading coefficient H and the message W. Let X_k denote the k^{th} coordinate of $f_H(W)$. In each time slot $k \in \{1, 2, \ldots, n\}$, s transmits X_k and d receives

$$Y_k = \sqrt{H} X_k + Z_k, \tag{10}$$

where $\{Z_k\}_{k=1}^n$ are i.i.d. standard normal random variables. In addition, we assume that (H, X_k) and Z_k are independent for each $k \in \{1, 2, \ldots, n\}$. In this work, we are interested in the *delay-limited* capacity [11] (also called the *zero-outage* capacity [12, Section 4.2.4]), which is the maximum transmission rate of a coding scheme under the constraint that the maximal error probability over all $H > 0$ (to be defined precisely in Definition 3) vanishes as n tends to infinity. The formal definition of the delay-limited capacity will be given later in Definition 5. Under the CSIT assumption and the long-term power constraint, a simple coding strategy that achieves the delay-limited capacity is s performing *channel inversion* [12, Section 4.2.4], i.e., using one codebook and transmitting the codewords multiplied by $\frac{1}{\sqrt{H}}$ so that the fading effect disappears from the point of view of d.

The formal definition of an (n, M_n, P)-code for the quasi-static fading channel is given below.

Definition 1: An (n, M_n, P)-code consists of the following:

1) A message set

$$\mathcal{W} \triangleq \{1, 2, \ldots, M_n\} \tag{11}$$

at node s. Message W is uniform on $\mathcal{W}$ and independent of the fading coefficient H.

2) An encoding function

$$f_h : \mathcal{W} \to \mathbb{R}^n \tag{12}$$

for each $h \geq 0$, where f_h is the encoding function at node s for encoding W when the fading coefficient H is equal to h such that

$$X^n = f_H(W). \tag{13}$$

The h*-fading codebook* is defined to be $\{f_h(w) \,|\, w \in \mathcal{W}\}$. The set of h-fading codebooks should satisfy the following long-term power constraint:

$$\frac{1}{M_n} \sum_{w \in \mathcal{W}} \mathrm{E}_{p_H}\left[\|f_H(w)\|^2\right] \leq nP. \tag{14}$$

3) A decoding function

$$\varphi : \mathbb{R}^n \to \mathcal{W}, \tag{15}$$

where φ is the decoding function for W at node d such that

$$\hat{W} = \varphi(Y^n). \tag{16}$$

We now state the definition of the quasi-static fading channel and define the AWGN non-fading channel as a special case of the fading channel.

Definition 2: A *quasi-static fading channel* is characterized by the fading distribution p_H and the conditional probability density distribution $q_{Y|X,H}$ satisfying

$$q_{Y|H,X}(y|h,x) = \mathcal{N}(y - \sqrt{h}x\,; 0, 1) \tag{17}$$

such that the following holds for any (n, M_n, P)-code:

$$p_{H,W,X^n,Y^n}(h, w, x^n, y^n) = p_{H,W,X^n}(h, w, x^n) \prod_{k=1}^{n} p_{Y_k|H,X_k}(y_k|h, x_k) \tag{18}$$

for all $h \geq 0$, $w \in \mathcal{W}$, $x^n \in \mathbb{R}^n$ and $y^n \in \mathbb{R}^n$ where

$$p_{Y_k|H,X_k}(y_k|h, x_k) = q_{Y|H,X}(y_k|h, x_k). \tag{19}$$

The quasi-static fading channel is also called the *AWGN non-fading channel* if H is deterministically equal to 1.

Since $p_{Y_k|H,X_k}(y_k|h,x_k)$ does not depend on k for each fixed $h \geq 0$ by (19), it follows from (18) and (19) that the quasi-static fading channel is stationary conditioned on the channel state h in the sense that

$$p_{Y_k|H,W,X^k,Y^{k-1}}(y_k|h,w,x^k,y^{k-1}) = q_{Y|H,X}(y_k|h,x_k) \tag{20}$$

for all (h,w,x^k,y^{k-1}) such that $p_{H,W,X^k,Y^{k-1}}(h,w,x^k,y^{k-1}) > 0$. For any (n,M_n,P)-code defined on the quasi-static fading channel, let $p_{H,W,X^n,Y^n,\hat{W}}$ be the joint distribution induced by both the channel and the code. We can factorize $p_{H,W,X^n,Y^n,\hat{W}}$ as follows:

$$p_{H,W,X^n,Y^n,\hat{W}} \stackrel{\text{(a)}}{=} p_{H,W,X^n,Y^n} p_{\hat{W}|Y^n} \tag{21}$$

$$\stackrel{(18)}{=} p_{H,W,X^n} \left(\prod_{k=1}^{n} p_{Y_k|H,X_k}\right) p_{\hat{W}|Y^n} \tag{22}$$

$$\stackrel{\text{(b)}}{=} p_H p_W p_{X^n|W,H} \left(\prod_{k=1}^{n} p_{Y_k|H,X_k}\right) p_{\hat{W}|Y^n} \tag{23}$$

where

(a) follows from the fact $\hat{W}$ is a function of Y^n by Definition 1.
(b) follows from the fact that W and H are independent.

For the AWGN non-fading channel, since H is deterministically equal to 1, equation (23) reduces to

$$p_{W,X^n,Y^n,\hat{W}} = p_W p_{X^n|W} \left(\prod_{k=1}^{n} p_{Y_k|X_k}\right) p_{\hat{W}|Y^n} \tag{24}$$

where

$$p_{Y_k|X_k}(y_k|x_k) \stackrel{(19)}{=} q_{Y|H,X}(y_k|1,x_k) \tag{25}$$

for all $(x_k,y_k) \in \mathbb{R}^2$. We define the delay-limited capacity via the following three definitions.

Definition 3: For an (n,M_n,P)-code, we can calculate according to (23) the *error probability for delay-limited decoding* defined as

$$\sup_{h>0} \Pr\{\hat{W} \neq W | H = h\}. \tag{26}$$

We call an (n,M_n,P)-code with error probability for delay-limited decoding no larger than ε an (n,M_n,P,ε)-code.

Remark 1: If we view each positive realization of H as a slow fading (time-invariant fading) process, then the definition of error probability for delay-limited decoding in Definition 3 is equivalent to the error probability criterion used in Definition 2.2 in [11].

Remark 2: The error probability for delay-limited decoding defined in Definition 3 considers the supremum of error probabilities over all positive realizations of H. Therefore, for any (n,M_n,P,ε)-code, the error probability of delay-limited decoding is guaranteed to be less than ε regardless of fading, and hence fading does not contribute to an "outage event". In contrast, under the average error probability formalism which defines the error probability to be $\Pr\{\hat{W} \neq W\}$ (as in [9, Section II]), the source may choose to create an "outage event" by transmitting nothing when the fading coefficient falls below certain threshold while keeping the average error probability less than ε. As explained in [11, Section I], the error probability of delay-limited decoding compared to the average error probability is a more relevant error measure for delay-sensitive applications such as voice and video if the fading is so slow in the timescale of tolerable delays that outage events are costly.

Definition 4: A rate R is *achievable* for the fading channel if there exists a sequence of (n,M_n,P,ε_n)-codes such that

$$\liminf_{n\to\infty} \frac{1}{n} \log M_n \geq R \tag{27}$$

and

$$\lim_{n\to\infty} \varepsilon_n = 0. \tag{28}$$

Definition 5: The delay-limited capacity, denoted by C, is defined to be

$$C \triangleq \sup\{R \,|\, R \text{ is achievable}\}. \tag{29}$$

Note that the definition of the delay-limited capacity C in Definition 5 coincides with Definition 2.2 in [11]. Define

$$P_{\text{D-L}} \triangleq \frac{P}{\mathrm{E}_{p_H}[1/H]}\,, \tag{30}$$

which is positive and finite by (7). It is well-known (e.g., [11, Section III-B]) that

$$C = \frac{1}{2}\log(1+P_{\text{D-L}}), \tag{31}$$

which justifies the following definition of capacity-achieving codes for the quasi-static fading channel.

Definition 6: A sequence of $(n, M_n, P, \varepsilon_n)$-codes is said to be *capacity-achieving* if

$$\lim_{n\to\infty}\frac{1}{n}\log M_n = \frac{1}{2}\log(1+P_{\text{D-L}}) \tag{32}$$

and

$$\lim_{n\to\infty}\varepsilon_n = 0. \tag{33}$$

We are ready to present our main result, and its proof will be presented in Section III-B.

Theorem 1: Fix a sequence of $(n, M_n, P, \varepsilon_n)$-codes. For each $n \in \mathbb{N}$, define

$$p^*_{Y^n}(y^n) \triangleq \prod_{k=1}^{n}\mathcal{N}(y_k; 0, 1+P_{\text{D-L}}) \tag{34}$$

to be the product of the capacity-achieving output distribution, and let p_{Y^n} be the output distribution induced by the $(n, M_n, P, \varepsilon_n)$-code on the quasi-static fading channel. If the sequence of codes is capacity-achieving, then

$$\lim_{n\to\infty}\frac{1}{n}D(p_{Y^n}\|p^*_{Y^n}) = 0. \tag{35}$$

Remark 3: It was shown in [3, Section V] that (35) holds for the AWGN non-fading channel. Here, we strengthen the result in [3, Section V] by showing that (35) holds for the more general quasi-static fading channel.

III. Proof of Theorem 1

A. Preliminary Results for Proving Theorem 1

We first state the following simple inequality, whose proof is standard and therefore omitted. Readers who are interested in the proof may refer to the converse proof of the channel capacity of the AWGN non-fading channel in [13, Section 9.2].

Proposition 1: Fix an (n, M_n, P)-code on the quasi-static fading channel and let $p_{H,W,X^n,Y^n,\hat{W}}$ be the distribution induced by the (n, M_n, P)-code. Define for each $h \geq 0$ the average power of the h-fading codebook $\{f_h(w) \,|\, w \in \mathcal{W}\}$ as

$$S_h^{(n)} \triangleq \frac{1}{nM_n}\sum_{w\in\mathcal{W}}\|f_h(w)\|^2 \tag{36}$$

so that when the fading coefficient H is equal to $h \geq 0$, the average signal power received at d can be written as $hS_h^{(n)}$ (cf. (10)). Then for each $h \geq 0$,

$$I_{p_{X^n,Y^n|H=h}}(X^n; Y^n) \leq \frac{n}{2}\log(1+hS_h^{(n)}). \tag{37}$$

The proof of Theorem 1 relies on the following two lemmas. The proof of the following lemma is deferred to the Appendix because it involves technical arguments.

Lemma 2: Fix a sequence of (n, M_n, P)-codes. Define $S_h^{(n)}$ for each $n \in \mathbb{N}$ and each $h > 0$ as in (36) to be the average power of the h-fading codebook so that $hS_h^{(n)}$ represents the average signal power received at d when the fading coefficient H is equal to $h > 0$. If

$$\liminf_{n\to\infty} \left\{ \inf_{h>0} hS_h^{(n)} \right\} \geq P_{\text{D-L}}, \tag{38}$$

then

$$\lim_{n\to\infty} \mathrm{E}_{p_H}\left[HS_H^{(n)}\right] = P_{\text{D-L}}. \tag{39}$$

Lemma 2 provides a sufficient condition (38) for the convergence of the expected value of $HS_H^{(n)}$, which is the average signal power received at d. This convergence is extremely important for bounding the quantity

$$\limsup_{n\to\infty} \frac{1}{n} \int_{\mathbb{R}^+} p_H(h) D(p_{Y^n|X^n,H=h} \| p^*_{Y^n} | p_{X^n|H=h}) \mathrm{d}h \tag{40}$$

in the proofs of Theorems 1 and 3 (cf. (71) and (141)). The main challenge in proving Lemma 2 comes from the fact that the number of fading states is allowed to be uncountable. If the number of fading states is assumed to be finite, then the proof of Lemma 2 can be greatly simplified (see the explanation at the beginning of the Appendix). The following lemma relies on Lemma 2 and is the key to proving Theorem 1.

Lemma 3: Fix a sequence of $(n, M_n, P, \varepsilon_n)$-codes, and let $p_{H,W,X^n,Y^n,\hat{W}}$ be the joint distribution induced by the $(n, M_n, P, \varepsilon_n)$-code. Then for each $h > 0$ and each $n \in \mathbb{N}$,

$$(1-\varepsilon_n)M_n \leq 1 + I_{p_{W,Y^n|H=h}}(W;Y^n). \tag{41}$$

In addition, if the sequence of codes is capacity-achieving, then

$$\lim_{n\to\infty} \mathrm{E}_{p_H}\left[HS_H^{(n)}\right] = P_{\text{D-L}} \tag{42}$$

where $S_h^{(n)}$ is as defined in (36) for each $h > 0$.

Proof: Fix any sequence of $(n, M_n, P, \varepsilon_n)$-codes that is capacity-achieving. Let $p_{H,W,X^n,Y^n,\hat{W}}$ be the joint distribution induced by the $(n, M_n, P, \varepsilon_n)$-code. Since the sequence of codes is capacity-achieving, it follows from Definition 6 that

$$\lim_{n\to\infty} \frac{1}{n} \log M_n = \frac{1}{2}\log(1 + P_{\text{D-L}}) \tag{43}$$

and

$$\lim_{n\to\infty} \varepsilon_n = 0. \tag{44}$$

Fix an arbitrary $n \in \mathbb{N}$ and fix the corresponding $(n, M_n, P, \varepsilon_n)$-code. Then, (41) follows from the following chain of inequalities for each $h > 0$:

$$\begin{aligned}
\log M_n &\stackrel{\text{(a)}}{=} H_{p_W}(W) && (45)\\
&\stackrel{\text{(b)}}{=} H_{p_{W|H=h}}(W) && (46)\\
&= I_{p_{W,Y^n|H=h}}(W;Y^n) + H_{p_{W,Y^n|H=h}}(W|Y^n) && (47)\\
&\stackrel{\text{(c)}}{\leq} I_{p_{X^n,Y^n|H=h}}(X^n;Y^n) + H_{p_{W,Y^n|H=h}}(W|Y^n) && (48)\\
&\stackrel{\text{(d)}}{\leq} I_{p_{X^n,Y^n|H=h}}(X^n;Y^n) + 1 + \varepsilon_n \log M_n && (49)
\end{aligned}$$

where

(a) follows from the fact that W is uniform on $\mathcal{W}$.
(b) follows from the hypothesis that H and W are independent.
(c) uses the fact by (23) that

$$W \to X^n \to Y^n \tag{50}$$

forms a Markov chain when they are generated according to $p_{W,X^n,Y^n|H=h}$.

(d) follows from Fano's inequality and Definition 3.

In the rest of the proof, we will first show that

$$(1-\varepsilon_n)\log M_n \le 1 + \inf_{h>0}\left\{\frac{n}{2}\log(1+hS_h^{(n)})\right\}, \tag{51}$$

and then prove (42) by using (43) and (51). Given that $H=h$ for some $h>0$, since

$$Y^n = \sqrt{h}f_h(W) + Z^n \tag{52}$$

by (10) and

$$\mathrm{E}_{p_W}\left[\|\sqrt{h}f_h(W)\|^2\right] \overset{(36)}{=} nhS_h^{(n)}, \tag{53}$$

it follows from Proposition 1 that

$$I_{p_{X^n,Y^n|H=h}}(W;Y^n) \le \frac{n}{2}\log(1+hS_h^{(n)}), \tag{54}$$

which then implies from (49) that

$$(1-\varepsilon_n)\log M_n \le 1 + \frac{n}{2}\log(1+hS_h^{(n)}). \tag{55}$$

Since (55) holds for all $h>0$, we have (51). We are now ready to prove (42). Combining (43), (44) and (51), we have

$$\frac{1}{2}\log(1+P_{\text{D-L}}) \le \liminf_{n\to\infty}\left\{\inf_{h>0}\left\{\frac{1}{2}\log(1+hS_h^{(n)})\right\}\right\}, \tag{56}$$

which implies that

$$\liminf_{n\to\infty}\left\{\inf_{h>0} hS_h^{(n)}\right\} \ge P_{\text{D-L}}, \tag{57}$$

which then implies from Lemma 2 that (42) holds. ∎

B. *Proof of Theorem 1*

We are ready to prove Theorem 1. Suppose we are given a sequence of $(n, M_n, P, \varepsilon_n)$-codes which is capacity-achieving and hence satisfies

$$\lim_{n\to\infty}\frac{1}{n}\log M_n = \frac{1}{2}\log(1+P_{\text{D-L}}) \tag{58}$$

and

$$\lim_{n\to\infty}\varepsilon_n = 0 \tag{59}$$

by Definition 6. Let $p_{H,W,X^n,Y^n,\hat{W}}$ be the distribution induced by the (n, M_n, P, ε)-code. For each $n\in\mathbb{N}$, define $p^*_{Y^n}(y^n)$ to be the product of the capacity-achieving output distribution as in (34). In order to prove (35), we consider the following standard steps for each $n\in\mathbb{N}$:

$$\begin{aligned}
D(p_{Y^n}\|p^*_{Y^n}) &\le D(p_{H,Y^n}\|p_H p^*_{Y^n}) && (60)\\
&= D(p_{Y^n|H}\|p^*_{Y^n}|p_H) && (61)\\
&= \int_{\mathbb{R}^+} p_H(h) D(p_{Y^n|H=h}\|p^*_{Y^n})\mathrm{d}h && (62)\\
&= \int_{\mathbb{R}^+} p_H(h)\left(D(p_{Y^n|X^n,H=h}\|p^*_{Y^n}|p_{X^n|H=h}) - D(p_{Y^n|X^n,H=h}\|p_{Y^n|H=h}|p_{X^n|H=h})\right)\mathrm{d}h && (63)\\
&= \int_{\mathbb{R}^+} p_H(h)\left(D(p_{Y^n|X^n,H=h}\|p^*_{Y^n}|p_{X^n|H=h}) - I_{p_{X^n,Y^n|H=h}}(X^n;Y^n)\right)\mathrm{d}h. && (64)
\end{aligned}$$

Using (41) in Lemma 3 and (64), we have

$$D(p_{Y^n}\|p^*_{Y^n}) \le \int_{\mathbb{R}^+} p_H(h)\left(D(p_{Y^n|X^n,H=h}\|p^*_{Y^n}|p_{X^n|H=h}) - (1-\varepsilon_n)\log M_n + 1\right)\mathrm{d}h. \tag{65}$$

In order to obtain an upper bound on the divergence term in (65), we define $S_h^{(n)}$ as in (36) and consider the following chain of inequalities for each $h > 0$:

$$
\begin{aligned}
&D(p_{Y^n|X^n,H=h}\|p^*_{Y^n}|p_{X^n|H=h}) \\
&= \int_{\mathbb{R}^n} p_{X^n|H}(x^n|h)\int_{\mathbb{R}^n} p_{Y^n|X^n,H}(y^n|x^n,h)\log\frac{p_{Y^n|X^n,H}(y^n|x^n,h)}{p^*_{Y^n}(y^n)}\,\mathrm{d}y^n\mathrm{d}x^n &(66)\\
&\stackrel{\text{(a)}}{=} \int_{\mathbb{R}^n} p_{X^n|H}(x^n|h)\int_{\mathbb{R}^n} \mathcal{N}(y^n-\sqrt{h}x^n;0,1)\log\frac{\mathcal{N}(y^n-\sqrt{h}x^n;0,1)}{\mathcal{N}(y^n;0,1+P_{\text{D-L}})}\,\mathrm{d}y^n\mathrm{d}x^n &(67)\\
&\stackrel{\text{(b)}}{=} \int_{\mathbb{R}^n} p_{X^n|H}(x^n|h)\int_{\mathbb{R}^n} \mathcal{N}(z^n;0,1)\log\frac{\mathcal{N}(z^n;0,1)}{\mathcal{N}(z^n+\sqrt{h}x^n;0,1+P_{\text{D-L}})}\,\mathrm{d}z^n\mathrm{d}x^n &(68)\\
&\stackrel{(6)}{=} \frac{n}{2}\log(1+P_{\text{D-L}}) \\
&\qquad + \frac{1}{2(1+P_{\text{D-L}})}\int_{\mathbb{R}^n} p_{X^n|H}(x^n|h)\sum_{k=1}^{n}\int_{\mathbb{R}}\mathcal{N}(z_k;0,1)(-P_{\text{D-L}}z_k^2+2\sqrt{h}x_kz_k+hx_k^2)\mathrm{d}z_k\mathrm{d}x^n &(69)\\
&= \frac{n}{2}\log(1+P_{\text{D-L}}) + \frac{-nP_{\text{D-L}}+\int_{\mathbb{R}^n} p_{X^n|H}(x^n|h)\sum_{k=1}^n hx_k^2\,\mathrm{d}x^n}{2(1+P_{\text{D-L}})} &(70)\\
&\stackrel{\text{(c)}}{\le} \frac{n}{2}\log(1+P_{\text{D-L}}) + \frac{-nP_{\text{D-L}}+nhS_h^{(n)}}{2(1+P_{\text{D-L}})} &(71)
\end{aligned}
$$

where

(a) follows from Definition 2 and (34).
(b) follows by letting $z^n = y^n - \sqrt{h}x^n$.
(c) follows from the fact that

$$
\begin{aligned}
\int_{\mathbb{R}^n} p_{X^n|H}(x^n|h)\sum_{k=1}^n hx_k^2\,\mathrm{d}x^n &= \sum_{w\in\mathcal{W}} p_W(w)h\|f_h(w)\|^2 &(72)\\
&\stackrel{(36)}{=} nhS_h^{(n)}. &(73)
\end{aligned}
$$

Combining (65) and (71), we obtain that for all $n\in\mathbb{N}$

$$
\frac{1}{n}D(p_{Y^n}\|p^*_{Y^n}) \le \frac{1}{2}\log(1+P_{\text{D-L}}) - \frac{(1-\varepsilon_n)\log M_n}{n} + \frac{1}{n} + \frac{-P_{\text{D-L}}+\mathrm{E}_{p_H}[HS_H^{(n)}]}{2(1+P_{\text{D-L}})}\,. \tag{74}
$$

On the other hand,

$$
\limsup_{n\to\infty}\mathrm{E}_{p_H}[HS_H^{(n)}] = P_{\text{D-L}} \tag{75}
$$

by (58) and (42) in Lemma 3. Consequently,

$$
\begin{aligned}
\limsup_{n\to\infty}\frac{1}{n}D(p_{Y^n}\|p^*_{Y^n}) &\stackrel{\text{(a)}}{\le} \limsup_{n\to\infty}\left(\frac{1}{2}\log(1+P_{\text{D-L}}) - \frac{(1-\varepsilon_n)\log M_n}{n}\right) &(76)\\
&\stackrel{\text{(b)}}{=} 0, &(77)
\end{aligned}
$$

where

(a) follows from (74) and (75).
(b) follows from (58) and (59).

IV. Quasi-Static Fading Channel with Non-Vanishing Error Probabilities

In this section, we extend Theorem 1 to capacity-achieving codes with *non-vanishing* error probabilities.

A. Problem Formulation and Main Result

We are interested in the two formalisms of error probability presented in the following definition.

Definition 7: For an (n, M_n, P)-code, we can calculate according to (23) the *maximum-over-messages error for delay-limited decoding* defined as

$$\sup_{h>0} \max_{w\in\mathcal{W}} \Pr\{\hat{W} \neq w | W = w, H = h\}. \tag{78}$$

We call an (n, M_n, P)-code with maximum-over-messages error for delay-limited decoding no larger than ε an $(n, M_n, P, \varepsilon)_{\text{max}}$-code. Similarly, we can calculate the *average-over-messages error for delay-limited decoding* defined as

$$\sup_{h>0} \Pr\{\hat{W} \neq W | H = h\}, \tag{79}$$

where $\Pr\{\hat{W} \neq W | H = h\}$ for each $h > 0$ is averaged over the realizations of W. We call an (n, M_n, P)-code with average-over-messages error for delay-limited decoding no larger than ε an $(n, M_n, P, \varepsilon)_{\text{avg}}$-code.

Remark 4: We can see from Definition 7 that the maximum-over-messages error is a more stringent error criterion than the average-over-messages error. Therefore, any $(n, M_n, P, \varepsilon)_{\text{max}}$-code is also an $(n, M_n, P, \varepsilon)_{\text{avg}}$-code but not vice versa. The differentiation between the maximum-over-messages error and the average-over-messages error is necessary because as explained later in Remark 8, the two formalisms will yield different convergence results for capacity-achieving codes with non-vanishing error probabilities.

Definition 8: Let $\varepsilon \in (0, 1)$ be a real number. A rate R is ε*-achievable* for the fading channel if there exists a sequence of $(n, M_n, P, \varepsilon)_{\text{max}}$-codes such that

$$\liminf_{n\to\infty} \frac{1}{n} \log M_n \geq R. \tag{80}$$

Definition 9: Let $\varepsilon \in (0, 1)$ be a real number. The delay-limited ε-capacity for the fading channel, denoted by C_ε, is defined to be

$$C_\varepsilon \triangleq \sup\{R \,|\, R \text{ is } \varepsilon\text{-achievable}\}. \tag{81}$$

The first result in this section is the following multi-letter converse bound for the quasi-static fading channel, which will be proved in Section IV-B.

Theorem 2: Fix an $\varepsilon \in (0, 1)$ and a sequence of $(n, M_n, P, \varepsilon)_{\text{max}}$-codes, and let $p_{H,W,X^n,Y^n,\hat{W}}$ denote the probability distribution induced by the code. Then for each $h > 0$ and each $n \geq 8$, we have

$$\log M_n \leq I_{p_{X^n,Y^n|H=h}}(X^n; Y^n) + n^{\frac{2}{3}}(1 + hS_h^{(n)}) \left(\frac{7-\varepsilon}{1-\varepsilon}\right) + \log\left(\frac{4}{1-\varepsilon}\right). \tag{82}$$

In addition,

$$C_\varepsilon = \frac{1}{2}\log(1 + P_{\text{D-L}}) \,. \tag{83}$$

Remark 5: Since the delay-limited ε-capacity does not depend on ε by (83) in Theorem 2, the quasi-static fading channel possesses the strong converse property under the maximum-over-messages error formalism. To the best of our knowledge, this strong converse result was not known hitherto.

Remark 6: If the average-over-messages error (79) is used instead of the maximum-over-messages error (78), the strong converse property (83) in Theorem 2 ceases to hold. This follows from Theorem 77 in [14] that the ε-capacity under the average-over-messages error criterion is equal to $\frac{1}{2}\log\left(1 + \frac{P}{1-\varepsilon}\right)$ for the AWGN non-fading channel.

The following definition of capacity-achieving codes with non-vanishing error probabilities is justified by (83) in Theorem 2.

Definition 10: For any $\varepsilon \in (0, 1)$, a sequence of $(n, M_n, P, \varepsilon)_{\text{max}}$-codes is said to be *capacity-achieving* if

$$\lim_{n\to\infty} \frac{1}{n} \log M_n = \frac{1}{2}\log(1 + P_{\text{D-L}}). \tag{84}$$

The following theorem is the main result in this section, whose proof will be provided in Section IV-C.

Theorem 3: Fix an $\varepsilon \in (0,1)$ and a sequence of $(n, M_n, P, \varepsilon)_{\max}$-codes. For each $n \in \mathbb{N}$, let $p^*_{Y^n}(y^n)$ be the product of the capacity-achieving output distribution as defined in (34), and let p_{Y^n} be the output distribution induced by the $(n, M_n, P, \varepsilon)_{\max}$-code on the quasi-static fading channel. If the sequence of codes is capacity-achieving, then

$$\lim_{n\to\infty} \frac{1}{n} D(p_{Y^n} \| p^*_{Y^n}) = 0. \tag{85}$$

Remark 7: It was conjectured in [4, Remark 8] that (85) need not hold for the AWGN non-fading channel under the long-term power constraint (14). We show in Theorem 3 that (85) does hold for the AWGN non-fading channel as well as the more general quasi-static fading channel.

Remark 8: If the maximum-over-messages error (78) is replaced with the average-over-messages error (79), the convergence result (85) no longer holds. This can be seen from the counterexample suggested in [4, Remark 8].

B. Proof of Theorem 2

To prove Theorem 2, we first state the following result in [4].

Theorem 4: Fix any $\varepsilon \in (0,1)$, any $n \in \mathbb{N}$ and any $(n, M_n, P, \varepsilon)_{\max}$-code on the AWGN non-fading channel, and let $p_{W,X^n,Y^n,\hat{W}}$ denote the probability distribution induced by the code. Define

$$P^{(n)}_{\max} \triangleq \frac{1}{n} \max_{w\in\mathcal{W}} \|f_1(w)\|^2 \tag{86}$$

to be the peak power of the codewords. Then,

$$\log M_n \le I_{p_{X^n,Y^n}}(X^n;Y^n) + \sqrt{\frac{4(3P^{(n)}_{\max}+2)n}{1-\varepsilon}} + \log\left(\frac{2}{1-\varepsilon}\right). \tag{87}$$

Proof: The theorem follows by Corollary 4 and the inequality (119) in [4] by the identifications $P_{Y|X} \equiv p_{Y^n|X^n}$, $Q_Y \equiv p_{Y^n}$ and $P_X \equiv p_{X^n}$. ∎

Theorem 4 cannot be used directly for proving Theorem 2 because $P^{(n)}_{\max}$ can grow with n even though the 1-fading codebook $\{f_1(w) \,|\, w \in \mathcal{W}\}$ satisfies the long term power constraint (14). However, Theorem 4 can be used to prove the following multi-letter converse bound for the AWGN non-fading channel, which is the key to proving Theorem 2.

Lemma 4: Fix an $\varepsilon \in (0,1)$ and a natural number $n \ge 8$. Suppose we are given an $(n, M_n, P, \varepsilon)_{\max}$-code on the AWGN non-fading channel, and let $p_{W,X^n,Y^n,\hat{W}}$ be the distribution induced by the $(n, M_n, P, \varepsilon)_{\max}$-code. Then,

$$\log M_n \le I_{p_{X^n,Y^n}}(X^n;Y^n) + n^{\frac{2}{3}}(1+P)\left(\frac{7-\varepsilon}{1-\varepsilon}\right) + \log\left(\frac{4}{1-\varepsilon}\right) \tag{88}$$

Proof: Fix an $\varepsilon \in (0,1)$ and let $p_{W,X^n,Y^n,\hat{W}}$ be the distribution induced by the given $(n, M_n, P, \varepsilon)_{\max}$-code. Define

$$\mathcal{A} \triangleq \left\{ w \in \mathcal{W} \,\middle|\, \|f_1(w)\|^2 \le n^{4/3} P \right\} \tag{89}$$

such that all the codewords in $\{f_1(w)|w\in\mathcal{A}\}$ have power no greater than $n^{1/3}P$, i.e.,

$$\frac{1}{n}\|f_1(w)\|^2 \le n^{1/3}P. \tag{90}$$

Define $p_{E|W}$ to be the conditional distribution such that

$$p_{E|W}(e|w) \triangleq \begin{cases} 1 & \text{if } e = \mathbf{1}\{w\in\mathcal{A}\}, \\ 0 & \text{otherwise,} \end{cases} \tag{91}$$

and define

$$p_{W,E,X^n,Y^n,\hat{W}} \triangleq p_{W,X^n,Y^n,\hat{W}}\, p_{E|W} \tag{92}$$

such that the following two statements hold:

(i) The distribution induced by the $(n, M_n, P, \varepsilon)_{\max}$-code is a marginal distribution of $p_{W,E,X^n,Y^n,\hat{W}}$ and

$$p_{W,E,X^n,Y^n,\hat{W}} = p_{W,X^n,Y^n,\hat{W}} p_{E|W} \tag{93}$$

$$\stackrel{(24)}{=} p_{W,X^n} p_{E|W} \left(\prod_{k=1}^{n} p_{Y_k|X_k} \right) p_{\hat{W}|Y^n}. \tag{94}$$

(ii) (W, E) is distributed according to $p_{W,E}$ such that

$$E = \mathbf{1}\{W \in \mathcal{A}\}. \tag{95}$$

In order to obtain an upper bound on the probability of W falling outside $\mathcal{A}$, consider

$$\Pr_{p_W}\{W \notin \mathcal{A}\} \stackrel{(89)}{=} \Pr_{p_W}\left\{ \|f_1(W)\|^2 > n^{4/3} P \right\} \tag{96}$$

$$\stackrel{\text{(a)}}{\leq} \frac{\mathrm{E}_{p_W}\left[\|f_1(W)\|^2 \right]}{n^{4/3} P} \tag{97}$$

$$\stackrel{\text{(b)}}{\leq} \frac{1}{n^{1/3}} \tag{98}$$

where

(a) follows from Markov's inequality.

(b) follows from the long-term power constraint (14) and the fact that

$$p_W(w) = \frac{1}{M_n} \tag{99}$$

for each $w \in \mathcal{W}$ by Definition 1.

Since

$$|\mathcal{A}| \geq \left(1 - \frac{1}{n^{1/3}}\right) M_n \tag{100}$$

by (98) and $\mathcal{A}$ can be viewed as a subcodebook of the $(n, M_n, P, \varepsilon)_{\max}$-code, it follows from the definition of $\mathcal{A}$ in (89) and Theorem 4 that

$$\log M_n + \log\left(1 - \frac{1}{n^{1/3}}\right) \leq I_{p_{X^n,Y^n|E=1}}(X^n; Y^n) + \sqrt{\frac{4(3n^{1/3}P + 2)n}{1-\varepsilon}} + \log\left(\frac{2}{1-\varepsilon}\right) \tag{101}$$

$$\leq I_{p_{X^n,Y^n|E=1}}(X^n; Y^n) + 2n^{\frac{2}{3}} \sqrt{\frac{3P+2}{1-\varepsilon}} + \log\left(\frac{2}{1-\varepsilon}\right) \tag{102}$$

$$< I_{p_{X^n,Y^n|E=1}}(X^n; Y^n) + 2n^{\frac{2}{3}} \sqrt{\frac{3(1+P)}{1-\varepsilon}} + \log\left(\frac{2}{1-\varepsilon}\right) \tag{103}$$

$$< I_{p_{X^n,Y^n|E=1}}(X^n; Y^n) + \frac{6(1+P)n^{\frac{2}{3}}}{1-\varepsilon} + \log\left(\frac{2}{1-\varepsilon}\right). \tag{104}$$

Since $n \geq 8$, we have

$$1 - \frac{1}{n^{1/3}} \geq \frac{1}{2}, \tag{105}$$

which implies from (104) that

$$\log M_n \leq I_{p_{X^n,Y^n|E=1}}(X^n; Y^n) + \frac{6(1+P)n^{\frac{2}{3}}}{1-\varepsilon} + \log\left(\frac{4}{1-\varepsilon}\right). \tag{106}$$

Following (106), consider

$$p_E(1) I_{p_{X^n,Y^n|E=1}}(X^n; Y^n) = p_E(1) I_{p_{E,X^n,Y^n}}(X^n; Y^n | E = 1) \tag{107}$$

$$\leq I_{p_{E,X^n,Y^n}}(X^n; Y^n | E) \tag{108}$$

$$\leq H_{p_{Y^n}}(Y^n) - H_{p_{E,X^n,Y^n}}(Y^n|E, X^n) \tag{109}$$

$$\stackrel{\text{(a)}}{=} I_{p_{X^n,Y^n}}(X^n; Y^n) \tag{110}$$

where (a) follows from the fact due to (94) that

$$E \to X^n \to Y^n \tag{111}$$

forms a Markov chain. Combining (98), (106) and (110), we have

$$\log M_n \leq \frac{I_{p_{X^n,Y^n}}(X^n; Y^n)}{1 - n^{-1/3}} + \frac{6(1+P)n^{\frac{2}{3}}}{1-\varepsilon} + \log\left(\frac{4}{1-\varepsilon}\right) \tag{112}$$

$$= I_{p_{X^n,Y^n}}(X^n; Y^n) + \frac{n^{-1/3} I_{p_{X^n,Y^n}}(X^n; Y^n)}{1 - n^{-1/3}} + \frac{6(1+P)n^{\frac{2}{3}}}{1-\varepsilon} + \log\left(\frac{4}{1-\varepsilon}\right) \tag{113}$$

$$\stackrel{(105)}{\leq} I_{p_{X^n,Y^n}}(X^n; Y^n) + 2n^{-1/3} I_{p_{X^n,Y^n}}(X^n; Y^n) + \frac{6(1+P)n^{\frac{2}{3}}}{1-\varepsilon} + \log\left(\frac{4}{1-\varepsilon}\right) \tag{114}$$

$$\stackrel{\text{(a)}}{\leq} I_{p_{X^n,Y^n}}(X^n; Y^n) + n^{\frac{2}{3}}\left(\log(1+P) + \frac{6(1+P)}{1-\varepsilon}\right) + \log\left(\frac{4}{1-\varepsilon}\right) \tag{115}$$

$$< I_{p_{X^n,Y^n}}(X^n; Y^n) + n^{\frac{2}{3}}\left(1 + P + \frac{6(1+P)}{1-\varepsilon}\right) + \log\left(\frac{4}{1-\varepsilon}\right) \tag{116}$$

$$= I_{p_{X^n,Y^n}}(X^n; Y^n) + n^{\frac{2}{3}}(1+P)\left(\frac{7-\varepsilon}{1-\varepsilon}\right) + \log\left(\frac{4}{1-\varepsilon}\right) \tag{117}$$

where (a) follows from Proposition 1. ∎

We are now ready to prove Theorem 2.

Proof of Theorem 2: Fix an $\varepsilon \in (0,1)$, a natural number $n \geq 8$ and an $(n, M_n, P, \varepsilon)_{\max}$-code, and let $p_{H,W,X^n,Y^n,\hat{W}}$ denote the probability distribution induced by the code. By Definition 3, we have for each $h > 0$

$$\max_{w \in \mathcal{W}} \Pr\{\hat{W} \neq w \,|\, W = w, H = h\} \leq \varepsilon. \tag{118}$$

Define $S_h^{(n)}$ as in (36) to be the average power of the h-fading codebook $\{f_h(w) \,|\, w \in \mathcal{W}\}$ for each $h > 0$. For every $h > 0$, we have the following two observations:

(i) The codebook $\{\sqrt{h} f_h(w) \,|\, w \in \mathcal{W}\}$ satisfies the long-term power constraint

$$\frac{1}{M_n} \sum_{w \in \mathcal{W}} \|\sqrt{h} f_h(w)\|^2 \stackrel{(36)}{=} n h S_h^{(n)}. \tag{119}$$

(ii) The codebook $\{\sqrt{h} f_h(w) \,|\, w \in \mathcal{W}\}$ can be viewed as an $(n, M_n, hS_h^{(n)}, \varepsilon)_{\max}$-code for the AWGN non-fading channel because the h-fading codebook $\{f_h(w) \,|\, w \in \mathcal{W}\}$ forms an $(n, M_n, S_h^{(n)}, \varepsilon)_{\max}$-code for the fading channel with $H = h$ (cf. (10)).

Based on the above observations, we can apply Lemma 4 to the codebook $\{\sqrt{h} f_h(w) \,|\, w \in \mathcal{W}\}$ for each $h > 0$ and obtain (82).

It remains to prove (83). To this end, we use (82) to obtain

$$\log M_n \leq \inf_{h>0}\left\{\frac{n}{2}\log(1 + hS_h^{(n)}) + n^{\frac{2}{3}}(1 + hS_h^{(n)})\left(\frac{7-\varepsilon}{1-\varepsilon}\right)\right\} + \log\left(\frac{4}{1-\varepsilon}\right). \tag{120}$$

In the following, we would like to show

$$\inf_{h>0}\left\{\frac{n}{2}\log(1 + hS_h^{(n)}) + n^{\frac{2}{3}}(1 + hS_h^{(n)})\left(\frac{7-\varepsilon}{1-\varepsilon}\right)\right\} \leq \frac{n}{2}\log(1 + P_{\text{D-L}}) + n^{\frac{2}{3}}(1 + P_{\text{D-L}})\left(\frac{7-\varepsilon}{1-\varepsilon}\right) \tag{121}$$

by assuming the contrary, i.e., there exists some $\zeta > 0$ such that

$$\frac{n}{2}\log(1 + hS_h^{(n)}) + n^{\frac{2}{3}}(1 + hS_h^{(n)})\left(\frac{7-\varepsilon}{1-\varepsilon}\right) \geq \frac{n}{2}\log(1 + P_{\text{D-L}}) + n^{\frac{2}{3}}(1 + P_{\text{D-L}})\left(\frac{7-\varepsilon}{1-\varepsilon}\right) + \zeta \tag{122}$$

for all $h > 0$, which implies that

$$hS_h^{(n)} > P_{\text{D-L}} \tag{123}$$

for each $h > 0$, which then implies that

$$\mathrm{E}_{p_H}[S_H^{(n)}] > P_{\text{D-L}}\mathrm{E}_{p_H}[1/H]. \tag{124}$$

Since

$$\mathrm{E}_{p_H}[S_H^{(n)}] \le P \tag{125}$$

by the long-term power constraint (14) and the definition of $S_h^{(n)}$ in (36), it follows from (124) that

$$P > P_{\text{D-L}}\mathrm{E}_{p_H}[1/H], \tag{126}$$

which contradicts the definition of $P_{\text{D-L}}$ in (30). Consequently, the assumption (122) is incorrect and (121) holds. Using (120) and (121), we obtain

$$\log M_n \le \frac{n}{2}\log(1+P_{\text{D-L}}) + n^{\frac{2}{3}}(1+P_{\text{D-L}})\left(\frac{7-\varepsilon}{1-\varepsilon}\right) + \log\left(\frac{4}{1-\varepsilon}\right), \tag{127}$$

which implies from Definition 9 that

$$C_\varepsilon \le \frac{1}{2}\log(1+P_{\text{D-L}}). \tag{128}$$

On the other hand, it follows from combining the channel coding theorem for the AWGN non-fading channel and the channel inversion strategy [12, Section 4.2.4] that

$$C_\varepsilon \ge \frac{1}{2}\log(1+P_{\text{D-L}}). \tag{129}$$

Combining (128) and (129) yields (83). For the sake of completeness, we include the explanation of (129) below. By the channel coding theorem, there exists a sequence of $(n, M_n, P_{\text{D-L}}, \varepsilon)_{\max}$-codes on the AWGN non-fading channel such that

$$\lim_{n\to\infty}\frac{1}{n}\log M_n = \frac{1}{2}\log(1+P_{\text{D-L}}), \tag{130}$$

and we let $\{f_1(1), f_1(2), \ldots, f_1(M_n)\}$ denote the codebook (i.e., the 1-fading codebook) for the $(n, M_n, P_{\text{D-L}}, \varepsilon)_{\max}$-code. We now construct a sequence of codes on the fading channel based on the channel inversion strategy and the $(n, M_n, P_{\text{D-L}}, \varepsilon)_{\max}$-codes in the following way: For each $n \in \mathbb{N}$, construct a length-n code for the fading channel such that for each $h > 0$, the corresponding h-fading codebook $\{f_h(w)\,|\,w \in \mathcal{W}\}$ satisfies $|\mathcal{W}| = M_n$ and

$$f_h(w) = \frac{1}{\sqrt{h}}f_1(w) \tag{131}$$

for all $w \in \mathcal{W}$. If the fading coefficient H is equal to some $h > 0$, then s transmits $f_h(W) \overset{(131)}{=} \frac{1}{\sqrt{h}}f_1(W)$ so that the received symbol becomes

$$Y^n \overset{(10)}{=} f_1(W) + Z^n, \tag{132}$$

where Y^n does not depend on the realization of H. Consequently, the destination can use the decoder of the original $(n, M_n, P_{\text{D-L}}, \varepsilon)_{\max}$-code so that the maximum-over-messages error probability of the constructed code is upper bounded by ε regardless of the realization of H. In addition, the constructed code satisfies the long-term power constraint (14), because

$$\frac{1}{M_n}\sum_{w\in\mathcal{W}}\mathrm{E}_{p_H}\left[\|f_H(w)\|^2\right] = \frac{1}{M_n}\sum_{w\in\mathcal{W}}\int_{\mathbb{R}^+} p_H(h)\|f_h(w)\|^2\mathrm{d}h \tag{133}$$

$$\overset{(131)}{=} \frac{1}{M_n}\sum_{w\in\mathcal{W}}\int_{\mathbb{R}^+} p_H(h)\frac{1}{h}\|f_1(w)\|^2\mathrm{d}h \tag{134}$$

$$= \frac{1}{M_n}\sum_{w\in\mathcal{W}}\|f_1(w)\|^2\mathrm{E}_{p_H}[1/H] \tag{135}$$

$$\stackrel{(14)}{\leq} nP_{\text{D-L}}\mathrm{E}_{p_H}[1/H] \tag{136}$$

$$\stackrel{(30)}{=} nP. \tag{137}$$

Consequently, the sequence of constructed codes is a sequence of $(n, M_n, P, \varepsilon)_{\text{max}}$-codes on the fading channel by Definition 1. Since the constructed $(n, M_n, P, \varepsilon)_{\text{max}}$-codes satisfy (130), we have (129). ■

C. Proof of Theorem 3

Fix an $\varepsilon \in (0,1)$. Suppose we are given a sequence of $(n, M_n, P, \varepsilon)_{\text{max}}$-codes which is capacity-achieving and hence satisfies

$$\lim_{n\to\infty} \frac{1}{n}\log M_n = \frac{1}{2}\log(1+P_{\text{D-L}}) \tag{138}$$

by Definition 10. Let $p_{H,W,X^n,Y^n,\hat{W}}$ be the distribution induced by the $(n, M_n, P, \varepsilon)_{\text{max}}$-code. For each $n \in \mathbb{N}$, define $p^*_{Y^n}(y^n)$ to be the product of the capacity-achieving output distribution as in (34). In order to prove (85), we follow standard steps (as those leading to (64)) to obtain

$$D(p_{Y^n}\|p^*_{Y^n}) \leq \int_{\mathbb{R}^+} p_H(h)\left(D(p_{Y^n|X^n,H=h}\|p^*_{Y^n}|p_{X^n|H=h}) - I_{p_{X^n,Y^n|H=h}}(X^n;Y^n)\right)\mathrm{d}h \tag{139}$$

for all $n \in \mathbb{N}$. Applying Theorem 2 to (139), we obtain that for all $n \geq 8$

$$\begin{aligned} &D(p_{Y^n}\|p^*_{Y^n}) \\ &\leq \int_{\mathbb{R}^+} p_H(h)\left(D(p_{Y^n|X^n,H=h}\|p^*_{Y^n}|p_{X^n|H=h}) - \log M_n + n^{\frac{2}{3}}(1+hS_h^{(n)})\left(\frac{7-\varepsilon}{1-\varepsilon}\right) + \log\left(\frac{4}{1-\varepsilon}\right)\right)\mathrm{d}h. \end{aligned} \tag{140}$$

In order to obtain an upper bound on the divergence term in (140), we follow the steps that lead to (71) and obtain for each $h > 0$

$$D(p_{Y^n|X^n,H=h}\|p^*_{Y^n}|p_{X^n|H=h}) \leq \frac{n}{2}\log(1+P_{\text{D-L}}) + \frac{-nP_{\text{D-L}} + nhS_h^{(n)}}{2(1+P_{\text{D-L}})}. \tag{141}$$

Combining (140) and (141), we obtain that for all $n \geq 8$

$$\begin{aligned} &\frac{1}{n}D(p_{Y^n}\|p^*_{Y^n}) \\ &\leq \frac{1}{2}\log(1+P_{\text{D-L}}) - \frac{1}{n}\log M_n + \frac{-P_{\text{D-L}} + \mathrm{E}_{p_H}[HS_H^{(n)}]}{2(1+P_{\text{D-L}})} + \frac{1}{n^{1/3}}(1+\mathrm{E}_{p_H}[HS_H^{(n)}])\left(\frac{7-\varepsilon}{1-\varepsilon}\right) + \frac{1}{n}\log\left(\frac{4}{1-\varepsilon}\right). \end{aligned} \tag{142}$$

It remains to show that

$$\limsup_{n\to\infty} \mathrm{E}_{p_H}[HS_H^{(n)}] \leq P_{\text{D-L}}, \tag{143}$$

and (85) will follow from (142) and (143). To this end, we obtain from Theorem 2 and Proposition 1 that for all $n \geq 8$,

$$\log M_n \leq \frac{n}{2}\log(1+hS_h^{(n)}) + n^{\frac{2}{3}}(1+hS_h^{(n)})\left(\frac{7-\varepsilon}{1-\varepsilon}\right) + \log\left(\frac{4}{1-\varepsilon}\right) \tag{144}$$

for all $h > 0$, which then implies that

$$\log M_n \leq \inf_{h>0}\left\{\frac{n}{2}\log(1+hS_h^{(n)}) + n^{\frac{2}{3}}(1+hS_h^{(n)})\left(\frac{7-\varepsilon}{1-\varepsilon}\right)\right\} + \log\left(\frac{4}{1-\varepsilon}\right). \tag{145}$$

Combining (138) and (145), we obtain

$$\frac{1}{2}\log(1+P_{\text{D-L}}) \leq \liminf_{n\to\infty}\inf_{h>0}\left\{\frac{1}{2}\log(1+hS_h^{(n)}) + \frac{1}{n^{1/3}}(1+hS_h^{(n)})\left(\frac{7-\varepsilon}{1-\varepsilon}\right)\right\}, \tag{146}$$

which implies that

$$\liminf_{n\to\infty}\left\{\inf_{h>0} hS_h^{(n)}\right\} \geq P_{\text{D-L}}. \tag{147}$$

Consequently, it follows from (147) and Lemma 2 that (143) holds. Combining (142) and (143), we obtain (85).

Appendix
Proof of Lemma 2

In this section, we omit in some places the mentioning that subsets of the real line such as $\mathcal{J}$, $\mathcal{K}$, $\tilde{\mathcal{K}}$ and $\hat{\mathcal{K}}$ are Borel-measurable for notational simplicity. We need the following proposition before proving Lemma 2. Note that the following proposition is straightforward if the number of fading states is assumed to be finite. Indeed, the proof of the following proposition is non-trivial and technical.

Proposition 5: For any $t\in(0,1]$,

$$\inf_{\mathcal{J}\in\mathcal{B}(\mathbb{R})\,:\,\Pr_{p_H}\{H\in\mathcal{J}\}\geq t} \mathrm{E}_{p_H}\left[\frac{1}{H}\middle| H\in\mathcal{J}\right] > 0 \tag{148}$$

where $\mathcal{B}(\mathbb{R})$ is the Borel σ-algebra on $\mathbb{R}$.

Proof: Fix a $t\in(0,1]$. If $t=1$, it follows that

$$\begin{aligned}\inf_{\mathcal{J}\,:\,\Pr_{p_H}\{H\in\mathcal{J}\}\geq t} \mathrm{E}_{p_H}\left[\frac{1}{H}\middle| H\in\mathcal{J}\right] &= \inf_{\mathcal{J}\,:\,\Pr_{p_H}\{H\in\mathcal{J}\}=1} \mathrm{E}_{p_H}\left[\frac{1}{H}\middle| H\in\mathcal{J}\right] && (149)\\ &= \mathrm{E}_{p_H}\left[\frac{1}{H}\right] && (150)\\ &\stackrel{(7)}{>} 0, && (151)\end{aligned}$$

which implies (148). Therefore, we assume in the rest of the proof that

$$0<t<1. \tag{152}$$

Let

$$a \triangleq \sup\left\{h\geq 0 \,|\, \Pr_{p_H}\{H\geq h\}\geq t\right\}, \tag{153}$$

where we do not specify that a depends on t for simplicity. Since $t>0$, we have

$$a<\infty. \tag{154}$$

We want to show that $a>0$ by assuming the contrary, i.e.,

$$a=0. \tag{155}$$

Using (153) and (155), we have for all $n\in\mathbb{N}$

$$\Pr_{p_H}\{H\geq 1/n\}<t, \tag{156}$$

which implies that for all $n\in\mathbb{N}$

$$\Pr_{p_H}\{H< 1/n\}\geq 1-t. \tag{157}$$

Since

$$\Pr_{p_H}\{H=0\} = \lim_{n\to\infty}\Pr_{p_H}\{H<1/n\} \tag{158}$$

by continuity of measure, it follows from (157) that

$$\Pr_{p_H}\{H=0\}\geq 1-t \stackrel{(152)}{>} 0, \tag{159}$$

contradicting the assumption of p_H in (7). Consequently, we conclude that

$$a>0. \tag{160}$$

In order to prove (148), we consider the following two cases:

Case (i): $\Pr_{p_H}\{H > a\} = 0$:

Since $\Pr_{p_H}\{H \le a\} = 1$,

$$\mathrm{E}_{p_H}\left[\frac{1}{H}\middle| H \in \mathcal{J}\right] \ge \frac{1}{a} \tag{161}$$

for any $\mathcal{J}$ with $\Pr_{p_H}\{H \in \mathcal{J}\} \ge t > 0$, which together with (160) implies (148).

Case (ii): $\Pr_{p_H}\{H > a\} > 0$:

It follows from the definition of a in (153) that

$$\Pr_{p_H}\{H > a\} \le t\,. \tag{162}$$

In order to show (148), we first fix an arbitrary $\mathcal{J}$ such that

$$\Pr_{p_H}\{H \in \mathcal{J}\} \ge t. \tag{163}$$

Using (162) and (163), we have

$$\Pr_{p_H}\{H \in \mathcal{J}\} \ge \Pr_{p_H}\{H > a\}\,. \tag{164}$$

In addition, we claim that

$$\mathrm{E}_{p_H}\left[\frac{1}{H}\middle| H > a\right] = \inf_{\mathcal{K}}\left\{\mathrm{E}_{p_H}\left[\frac{1}{H}\middle| H \in \mathcal{K}\right] \middle| \Pr_{p_H}\{H \in \mathcal{K}\} = \Pr_{p_H}\{H > a\}\right\}. \tag{165}$$

Note that the infimum is achieved by the set $\mathcal{K} = (a, \infty)$. To prove the above claim, it suffices to show that for any other $\tilde{\mathcal{K}}$ with equal measure as (a,∞) with respect to p_H, i.e.,

$$\Pr_{p_H}\{H \in \tilde{\mathcal{K}}\} = \Pr_{p_H}\{H > a\}, \tag{166}$$

we must have

$$\mathrm{E}_{p_H}\left[\frac{1}{H}\middle| H \in \tilde{\mathcal{K}}\right] - \mathrm{E}_{p_H}\left[\frac{1}{H}\middle| H > a\right] \ge 0 \tag{167}$$

due to the monotonicity of $h \mapsto 1/h$. To this end, fix any $\tilde{\mathcal{K}}$ that satisfies (166) and let $\mathcal{A} \triangleq (a,\infty) \setminus \tilde{\mathcal{K}}$ and $\mathcal{B} \triangleq \tilde{\mathcal{K}} \cap [0,a]$. Then, we have

$$\begin{aligned}
\Pr_{p_H}\{H \in \mathcal{A}\} &= \Pr_{p_H}\{H \in (a,\infty) \setminus \tilde{\mathcal{K}}\} && (168)\\
&= \Pr_{p_H}\{H \in (a,\infty)\} - \Pr_{p_H}\{H \in (a,\infty) \cap \tilde{\mathcal{K}}\} && (169)\\
&\overset{(166)}{=} \Pr_{p_H}\{H \in \tilde{\mathcal{K}}\} - \Pr_{p_H}\{H \in (a,\infty) \cap \tilde{\mathcal{K}}\} && (170)\\
&= \Pr_{p_H}\{H \in \tilde{\mathcal{K}} \cap [0,a]\} && (171)\\
&= \Pr_{p_H}\{H \in \mathcal{B}\}. && (172)
\end{aligned}$$

Consider the following chain of inequalities:

$$\begin{aligned}
&\mathrm{E}_{p_H}\left[\frac{1}{H}\middle| H \in \tilde{\mathcal{K}}\right] - \mathrm{E}_{p_H}\left[\frac{1}{H}\middle| H > a\right]\\
&= \int_{\tilde{\mathcal{K}}} \frac{p_{H|H\in\tilde{\mathcal{K}}}(h)}{h}\mathrm{d}h - \int_{(a,\infty)} \frac{p_{H|H>a}(h)}{h}\mathrm{d}h && (173)\\
&\overset{(166)}{=} \frac{1}{\Pr_{p_H}\{H > a\}}\left(\int_{\tilde{\mathcal{K}}} \frac{p_H(h)}{h}\mathrm{d}h - \int_{(a,\infty)} \frac{p_H(h)}{h}\mathrm{d}h\right) && (174)\\
&= \frac{1}{\Pr_{p_H}\{H > a\}}\left(\int_{\tilde{\mathcal{K}}\cap[0,a]} \frac{p_H(h)}{h}\mathrm{d}h + \int_{\tilde{\mathcal{K}}\cap(a,\infty)} \frac{p_H(h)}{h}\mathrm{d}h - \int_{(a,\infty)\cap\tilde{\mathcal{K}}} \frac{p_H(h)}{h}\mathrm{d}h - \int_{(a,\infty)\setminus\tilde{\mathcal{K}}} \frac{p_H(h)}{h}\mathrm{d}h\right) && (175)
\end{aligned}$$

$$= \frac{1}{\Pr_{p_H}\{H > a\}} \left(\int_{\tilde{\mathcal{K}} \cap [0,a]} \frac{p_H(h)}{h} \mathrm{d}h - \int_{(a,\infty)\setminus \tilde{\mathcal{K}}} \frac{p_H(h)}{h} \mathrm{d}h \right) \tag{176}$$

$$\overset{\text{(a)}}{\geq} \frac{1}{a\Pr_{p_H}\{H > a\}} \left(\int_{\tilde{\mathcal{K}} \cap [0,a]} p_H(h) \mathrm{d}h - \int_{(a,\infty)\setminus \tilde{\mathcal{K}}} p_H(h) \mathrm{d}h \right) \tag{177}$$

$$\overset{(172)}{=} 0, \tag{178}$$

where (a) follows from the monotonicity of $h \mapsto 1/h$. Consequently, the claim in (165) follows from (166) and (178). On the other hand, for any $\mathcal{K}$ with $\Pr_{p_H}\{H \in \mathcal{K}\} > 0$ and any positive $\lambda \leq \Pr_{p_H}\{H \in \mathcal{K}\}$, we have

$$\mathrm{E}_{p_H}\left[\frac{1}{H} \middle| H \in \mathcal{K} \right] \geq \inf_{\hat{\mathcal{K}}} \left\{ \mathrm{E}_{p_H}\left[\frac{1}{H} \middle| H \in \hat{\mathcal{K}} \right] \middle| \Pr_{p_H}\{H \in \hat{\mathcal{K}}\} = \lambda \right\} \tag{179}$$

because due to the monotonicity of $h \mapsto 1/h$, we can always construct a $\hat{\mathcal{K}} \subseteq \mathcal{K}$ with $\Pr_{p_H}\{H \in \hat{\mathcal{K}}\} = \lambda$ by excluding an appropriate leftmost subset of $\mathcal{K}$ such that $\mathrm{E}_{p_H}\left[\frac{1}{H} \middle| H \in \mathcal{K} \right] \geq \mathrm{E}_{p_H}\left[\frac{1}{H} \middle| H \in \hat{\mathcal{K}} \right]$. We now consider

$$\mathrm{E}_{p_H}\left[\frac{1}{H} \middle| H \in \mathcal{J} \right] \geq \inf_{\mathcal{K}} \left\{ \mathrm{E}_{p_H}\left[\frac{1}{H} \middle| H \in \mathcal{K} \right] \middle| \Pr_{p_H}\{H \in \mathcal{K}\} = \Pr_{p_H}\{H \in \mathcal{J}\} \right\} \tag{180}$$

$$\overset{\text{(a)}}{\geq} \inf_{\hat{\mathcal{K}}} \left\{ \mathrm{E}_{p_H}\left[\frac{1}{H} \middle| H \in \hat{\mathcal{K}} \right] \middle| \Pr_{p_H}\{H \in \hat{\mathcal{K}}\} = \Pr_{p_H}\{H > a\} \right\} \tag{181}$$

$$\overset{(165)}{=} \mathrm{E}_{p_H}\left[\frac{1}{H} \middle| H > a \right] \tag{182}$$

where (a) follows from (164) and (179). Since (182) holds for any arbitrary $\mathcal{J}$ that satisfies (163), we have

$$\inf_{\mathcal{J}\,:\,\Pr_{p_H}\{H \in \mathcal{J}\} \geq t} \mathrm{E}_{p_H}\left[\frac{1}{H} \middle| H \in \mathcal{J} \right] \geq \mathrm{E}_{p_H}\left[\frac{1}{H} \middle| H > a \right], \tag{183}$$

which then implies from the assumption under this case and (154) that

$$\inf_{\mathcal{J}\,:\,\Pr_{p_H}\{H \in \mathcal{J}\} \geq t} \mathrm{E}_{p_H}\left[\frac{1}{H} \middle| H \in \mathcal{J} \right] > 0. \tag{184}$$

■

We are now ready to prove Lemma 2.

Proof of Lemma 2: Fix a sequence of (n, M_n, P)-codes. Suppose (38) holds for the sequence of codes, which implies that

$$\liminf_{n\to\infty} \mathrm{E}_{p_H}\left[HS_H^{(n)} \right] \geq P_{\text{D-L}}. \tag{185}$$

We will prove (39) by assuming the contrary, i.e., there exists some $\delta > 0$ such that

$$\limsup_{n\to\infty} \mathrm{E}_{p_H}\left[HS_H^{(n)} \right] > P_{\text{D-L}}(1 + 2\delta), \tag{186}$$

which implies that there exists some subsequence of $\{n\}_{n=1}^{\infty}$, denoted by $\{n_\ell\}_{\ell=1}^{\infty}$, such that for all $\ell \in \mathbb{N}$,

$$\mathrm{E}_{p_H}\left[HS_H^{(n)} \right] > P_{\text{D-L}}(1 + \delta), \tag{187}$$

which then implies that there exists some $t > 0$ such that for all sufficiently large ℓ,

$$\Pr_{p_H}\left\{ HS_H^{(n_\ell)} > P_{\text{D-L}}(1 + \delta) \right\} = t > 0. \tag{188}$$

Define

$$\upsilon(\delta, t) \triangleq \frac{\delta t P_{\text{D-L}} \inf_{\mathcal{J} \in \mathcal{B}(\mathbb{R}) : \Pr_{p_H}\{H \in \mathcal{J}\} \geq t} \left\{ \mathrm{E}_{p_H}\left[\frac{1}{H} \middle| H \in \mathcal{J} \right] \right\}}{2\mathrm{E}_{p_H}\left[\frac{1}{H} \right]}. \tag{189}$$

Using Proposition 5 and the facts that $\delta > 0$, $t > 0$, $P_{\text{D-L}} > 0$ and $\mathrm{E}_{p_H}\left[\frac{1}{H}\right] > 0$, we have

$$\upsilon(\delta, t) > 0. \tag{190}$$

Using (38) and (190), we have for all sufficiently large ℓ

$$\inf_{h>0}\left\{hS_h^{(n_\ell)}\right\} > P_{\text{D-L}} - \upsilon(\delta, t). \tag{191}$$

Consider the following chain of inequalities for all sufficiently large ℓ:

$$\begin{aligned}
P &\overset{\text{(a)}}{\geq} \mathrm{E}_{p_H}\left[S_H^{(n_\ell)}\right] && (192)\\
&= \mathrm{E}_{p_H}\left[S_H^{(n_\ell)}\middle| HS_H^{(n_\ell)} \leq P_{\text{D-L}}(1+\delta)\right]\Pr\left\{HS_H^{(n_\ell)} \leq P_{\text{D-L}}(1+\delta)\right\}\\
&\quad + \mathrm{E}_{p_H}\left[S_H^{(n_\ell)}\middle| HS_H^{(n_\ell)} > P_{\text{D-L}}(1+\delta)\right]\Pr\left\{HS_H^{(n_\ell)} > P_{\text{D-L}}(1+\delta)\right\} && (193)\\
&\overset{(191)}{\geq} \mathrm{E}_{p_H}\left[\frac{P_{\text{D-L}} - \upsilon(\delta,t)}{H}\middle| HS_H^{(n_\ell)} \leq P_{\text{D-L}}(1+\delta)\right]\Pr\left\{HS_H^{(n_\ell)} \leq P_{\text{D-L}}(1+\delta)\right\}\\
&\quad + \mathrm{E}_{p_H}\left[S_H^{(n_\ell)}\middle| HS_H^{(n_\ell)} > P_{\text{D-L}}(1+\delta)\right]\Pr\left\{HS_H^{(n_\ell)} > P_{\text{D-L}}(1+\delta)\right\} && (194)\\
&\overset{\text{(b)}}{\geq} \mathrm{E}_{p_H}\left[\frac{P_{\text{D-L}} - \upsilon(\delta,t)}{H}\middle| HS_H^{(n_\ell)} \leq P_{\text{D-L}}(1+\delta)\right]\Pr\left\{HS_H^{(n_\ell)} \leq P_{\text{D-L}}(1+\delta)\right\}\\
&\quad + \mathrm{E}_{p_H}\left[\frac{P_{\text{D-L}}(1+\delta)}{H}\middle| HS_H^{(n_\ell)} > P_{\text{D-L}}(1+\delta)\right]\Pr\left\{HS_H^{(n_\ell)} > P_{\text{D-L}}(1+\delta)\right\} && (195)\\
&= \left(P_{\text{D-L}} - \upsilon(\delta,t)\right)\mathrm{E}_{p_H}\left[\frac{1}{H}\right]\\
&\quad + \left(P_{\text{D-L}}(\delta + \upsilon(\delta,t))\right)\mathrm{E}_{p_H}\left[\frac{1}{H}\middle| HS_H^{(n_\ell)} > P_{\text{D-L}}(1+\delta)\right]\Pr\left\{HS_H^{(n_\ell)} > P_{\text{D-L}}(1+\delta)\right\} && (196)\\
&\overset{(188)}{\geq} \left(P_{\text{D-L}} - \upsilon(\delta,t)\right)\mathrm{E}_{p_H}\left[\frac{1}{H}\right] + \delta t P_{\text{D-L}} \inf_{\mathcal{J}:\Pr_{p_H}\{H\in\mathcal{J}\}\geq t}\left\{\mathrm{E}_{p_H}\left[\frac{1}{H}\middle| H \in \mathcal{J}\right]\right\} && (197)\\
&\overset{(189)}{=} \left(P_{\text{D-L}} - \upsilon(\delta,t)\right)\mathrm{E}_{p_H}\left[\frac{1}{H}\right] + 2\upsilon(\delta,t)\,\mathrm{E}_{p_H}\left[\frac{1}{H}\right] && (198)\\
&= P_{\text{D-L}}\mathrm{E}_{p_H}\left[\frac{1}{H}\right] + \upsilon(\delta,t)\,\mathrm{E}_{p_H}\left[\frac{1}{H}\right] && (199)\\
&\overset{\text{(c)}}{>} P_{\text{D-L}}\mathrm{E}_{p_H}\left[\frac{1}{H}\right] && (200)\\
&\overset{(30)}{=} P && (201)
\end{aligned}$$

where

(a) follows from the long-term power constraint (14) and the definition of $S_h^{(n_\ell)}$ in (36).
(b) follows from the inequality in the second conditional expectation.
(c) follows from (190) and the fact that $\mathrm{E}_{p_H}\left[\frac{1}{H}\right] > 0$ by (7).

Since (201) is a contradiction due to the strict inequality, the assumption in (186) does not hold for all $\delta > 0$, which implies that

$$\limsup_{n\to\infty} \mathrm{E}_{p_H}\left[HS_H^{(n)}\right] \leq P_{\text{D-L}}. \tag{202}$$

Combining (185) and (202), we obtain (39). ∎

Acknowledgments

The authors are supported by an NUS grant (R-263-000-A98-750/133) an NUS Young Investigator Award (R-263-000-B37-133), and a Ministry of Education (MOE) Tier 2 grant (R-263-000-B61-112).

References

[1] T. S. Han and S. Verdú, "Approximation of output statistics," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 752–772, 1993.

[2] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147–1157, 1994.

[3] S. Shamai and S. Verdú, "The empirical distribution of good codes," *IEEE Trans. Inf. Theory*, vol. 43, no. 3, pp. 836–846, 1997.

[4] Y. Polyanskiy and S. Verdú, "Empirical distribution of good channel codes with nonvanishing error probability," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 5–21, 2014.

[5] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.

[6] M. Hayashi, "Information spectrum approach to second-order coding rate in channel coding," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 4947–4966, 2009.

[7] V. Y. F. Tan, "Asymptotic estimates in information theory with non-vanishing error probabilities," *Foundations and Trends® in Communications and Information Theory*, vol. 11, no. 1–2, 2014.

[8] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1468–1489, 1999.

[9] W. Yang, G. Caire, G. Durisi, and Y. Polyanskiy, "Optimum power control at finite blocklength," *IEEE Trans. Inf. Theory*, vol. 61, no. 9, pp. 4598–4615, 2015.

[10] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge University Press, 2005.

[11] S. V. Hanly and D. N. C. Tse, "Multiaccess fading channels – Part II: Delay-limited capacities," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2816–2831, 1998.

[12] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge University Press, 2005.

[13] T. M. Cover, *Elements of Information Theory*, 2nd ed. Wiley, 2006.

[14] Y. Polyanskiy, "Channel coding: Non-asymptotic fundamental limits," Ph.D. dissertation, Princeton University, 2010.